EpCAM^{+}CD73^{+} mark epithelial progenitor cells in postnatal human lung and is associated with pathogenesis of pulmonary disease including lung adenocarcinoma

Limei Wang[1,2], Patrick Dorn[1], Cedric Simillion[3], Laurène Froment[1,2], Sabina Berezowska[4], Stefan A. Tschanz[5], Beat Haenni[5], Fabian Blank[2,6], Carlos Wotzkow[6], Ren-Wang Peng[1,2], Thomas M. Marti[1,2], Peter K. Bode[7], Ueli Moehrlen[8], Ralph A. Schmid[1,2†] and Sean R.R. Hall[1,2†]

[1]Division of General Thoracic Surgery, Inselspital, Bern University Hospital, Bern, Switzerland
[2]Department of BioMedical Research, University of Bern, Switzerland
[3]Interfaculty Bioinformatics, University of Bern, Switzerland
[4]Institute of Pathology, University of Bern, Bern, Switzerland
[5]Institute of Anatomy, University of Bern, Switzerland
[6]DCR Live Imaging Core, University of Bern, Switzerland
[7]Department of Pathology and Molecular Pathology, University Hospital Zurich, Switzerland
[8]Department of Pediatric Surgery, University Children's Hospital, Zurich, Switzerland

†Address correspondence to Sean R.R. Hall or Ralph A. Schmid, [1]Division of General Thoracic Surgery, Inselspital, Bern University Hospital and Department of BioMedical Research, University of Bern, Murtenstrasse 50, 3008 Bern, Switzerland
Tel. +41 031 632 2300
Fax. +41 031 632 2327
Email: Sean.Hall@insel.ch or Ralph.Schmid@insel.ch

Funding: LW is a doctoral student supported by a 4-year China Scholarship Council award. Laser scanning microscopy imaging was funded by the R'Equip grant from the Swiss National Science Foundation Nr. 316030_145003.

Running Title: CD73 progenitor cells in human lung.

Abstract word count: 250

Lung injury in mice induces mobilization of discrete subsets of epithelial progenitor cells to promote new airway and alveolar structures. However, whether similar cell types exist in human lung remains unresolved. Using flow cytometry, we identified a distinct cluster of cells expressing epithelial cell adhesion molecule (EpCAM), a cell surface marker expressed on epithelial progenitor cells, enriched in the ecto-5'-nucleotidase CD73 in unaffected postnatal human lung resected from pediatric patients with congenital lung lesions. Within the $EpCAM^{+}CD73^{+}$ population, a small subset co-express integrin β4 and HTII-280. This population remained stable with age. Spatially, $EpCAM^{+}CD73^{+}$ cells were positioned along the basal membrane of respiratory epithelium and alveolus next to $CD73^{+}$ cells lacking EpCAM. Expanded $EpCAM^{+}CD73^{+}$ cells give rise to pseudostratified epithelium in 2D air-liquid interface or a clonal 3D organoid assay. Organoids generated under alveolar differentiation conditions were cystic-like and lacked robust alveolar mature cell types. Compared with unaffected postnatal lung, congenital lung lesions were marked by clusters of $EpCAM^{+}CD73^{+}$ cells in airway and cystic distal lung structures lined by simple epithelium of composed of $EpCAM^{+}SCGB1A1^{+}$ cells and hyperplastic $EpCAM^{+}proSPC^{+}$ cells. In non-small cell lung cancer (NSCLC), there was a marked increase in $EpCAM^{+}CD73^{+}$ tumor cells enriched in inhibitory immune checkpoint molecules CD47 and programmed death-ligand 1 (PD-L1), which was associated with poor survival in lung adenocarcinoma. In conclusion, $EpCAM^{+}CD73^{+}$ cells are a rare novel epithelial progenitor cell in human lung. Importantly, re-emergence of CD73 in lung adenocarcinoma enriched in negative immune checkpoint molecules may serve as a novel therapeutic target.

Keywords: EpCAM; congenital lung lesions; organoids; CD73; immune checkpoint; adenocarcinoma

Introduction

Despite being a quiescent organ at homeostasis, clinically the human lung does possess an ability for repair after various insults, although inherently low (5, 45). Nonetheless, the cellular and molecular mechanisms governing the regenerative process following lung injury have been slow in forthcoming. Moreover, the idea of a resident cell type within human lung with facultative stem cell function that is induced following injury remains controversial (25). Identifying putative cell populations in the human adult lung with facultative function and whether certain disease settings affect their numbers and function, may uncover new targets for therapeutic treatment to restore normal lung structure and function following various injuries.

In human lung, epithelial cell adhesion molecule (EpCAM/CD326) is used as a biomarker enriching a population of cells endowed with stem/progenitor-like function (2). This type I single span transmembrane glycoprotein was first described as a cell surface antigen on human carcinoma cells of epithelial origin (17). Since then, a growing body of evidence has shown that EpCAM is expressed on a dynamic range of cells and is critically involved in ensuring proper endodermal/epithelial morphogenesis (46). During development, EpCAM expression wanes in terminally differentiated cells and re-emerges during tissue regeneration and malignancy (34). Recently, a rare population of EpCAM-positive cells enriched in ecto-5'-nucleotidase/CD73 in human breast tissue was shown to possess enhanced cell plasticity giving rise to tissue from all three germ layers (38). Following this, high-dimensional analysis of single cells during cellular reprogramming revealed that $CD73^{high}Ki67^{High}$ distinguishes partially reprogrammed cells that were $Oct4^{high}Klf4^{high}EpCAM^{low}$ (52). This transitional cell state was found to precede mesenchymal-to-epithelial transition and then pluripotency (52). CD73 is a cell

surface ectoenzyme that hydrolyzes the conversion of extracellular adenosine monophosphate to adenosine (33). In the human lung, CD73 is expressed by a wide number of cells within the mesenchymal and immune compartments (33). Within the epithelial compartment, CD73 is locally expressed on the airway mucosal and serosal surface where it is functionally active in the conversion of extracellular ATP to adenosine under normal physiological conditions (37). Hypoxia results in upregulation in CD73/adenosine, which under chronic conditions becomes maladaptive (3). Overexpression of cell surface CD73 is associated with worse clinical outcome linked with excess adenosine production in both breast (4) and ovarian cancer (47). Whether CD73 can be used to identify a discrete population of $EpCAM^+$ cells with progenitor-like function in human lung is currently not known. Furthermore, the functional significance of this population in various settings of lung injury remain unexplored.

Here, using a multiparametric approach we demonstrate that $EpCAM^+CD73^+$ cells represent a rare but stable progenitor-like population in human lung. $EpCAM^+CD73^+$ are positioned along the basal membrane of respiratory epithelium, as well as in the alveolar region in unaffected postnatal human lung. Single $EpCAM^+CD73^+$ cells give rise to pseudostratified epithelium in two-dimensional air-liquid interface or in a clonal 3D organoid assay, whereas under alveolar differentiation conditions clonal organoids were more cystic-like and lacked robust alveolar mature cell types. Analysis of congenital lung lesions revealed the presence of clusters of $EpCAM^+CD73^+$ cells in the airway, whereas cystic distal lung structures were lined by simple epithelium consisting of $EpCAM^+SCGB1A1^+$ cells and hyperplastic $EpCAM^+proSPC^+$ cells. An upregulation of $EpCAM^+CD73^+$ tumor cells enriched in inhibitory immune checkpoint molecules CD47

and programmed death-ligand 1 (PD-L1), which was associated with poor survival, was revealed in lung adenocarcinoma.

Materials and Methods

Collection of lung tissue

Collection and processing of human lung tissue was performed as previously described (49). Briefly, resected tissues were collected from pediatric patients undergoing elective surgery for congenital lung lesions and other lung abnormalities at Children's University Hospital of Zurich (see Table S1), as well as adult patients undergoing resection for non-small cell lung cancer (NSCLC) or metastasis to the lung (see Table S2 and S3) at Bern University Hospital, Department of Thoracic Surgery. The use of surgically resected material for research purposes was provided by all patients included in this study, which was approved by the Ethics Commission of the Canton of Bern (KEK-BE:042/2015). All supplemental material is available at https://doi.org/10.6084/m9.figshare.12488867.

Flow cytometric analysis, prospective cell isolation and generation of primary cultures

Generation of single cell suspensions for flow cytometric analysis and prospective cell isolation were performed as previously described (49). Briefly, single cells from digested lung tissue were stained with fluorescently conjugated human monoclonal antibodies targeting: CD45, CD14, CD31, CD73, CD90, EpCAM (see Table S7). Following sorting, $EpCAM^{+}CD73^{+}$ cells (see Figure 1C and Supplemental Fig. S1 for full gating strategy) were plated on dishes pre-coated with a solution consisting of 0.2% gelatin, 0.3 mg/ml of human collagen I (Sigma) and 0.03 mg/ml of human collagen IV (Sigma). Cells were grown in cell expansion media (see Table S8) and maintained in a humidified 37°C low oxygen (3% O_2) incubator in 5% CO_2. An additional FACS analysis was performed on a second cohort of postnatal (n = 6; PL017-PL022) and adult lungs (n = 6; patient 10-15,

see Table S1) using the same antibody backbone as described above with addition of the following antibodies: CD146, NOTCH3 and integrin β4 (CD104) (see Table S7). To further immunophenotype the epithelial progenitor cell population, we stained a third cohort of lung tissue (postnatal, n = 6; PL029-PL035 and adult lungs, n = 7; patient, see Table S1 and S2) using our original antibody backbone with addition of the following antibodies: HTII-280, CD24 and PDPN (see Table S7).

Immunophenotype of cell subsets using flow cytometry

Following isolation and expansion of $EpCAM^{+}CD73^{+}$ cells (see Figure 2A and Table S10), cells were harvested and re-suspended in FACS staining buffer. Following Fc block, cells were incubated with the following fluorescently conjugated human monoclonal antibodies: integrin β4, EpCAM, CD47 (see Table S7). Cells were incubated on ice in the dark for 30 minutes. To exclude dead cells and debris, 7-AAD was added. Cell acquisition was performed using a BD FACS LSRII. For analysis, a minimum of 10,000 events were collected and analyzed using FlowJo software version 10.7 (Tree Star).

Immunofluorescence analysis

Preparation of lung tissue for immunostaining was performed as previously described (49). Briefly, 5 µm sections of were stained with hematoxylin & eosin using standard protocols. For immunofluorescence, 5 µm sections were stained with a panel of human monoclonal antibodies: EpCAM, CD73, CD90, SOX2, KRT5, TRP63, integrin β4, SCGB1A1 and proSPC on different tissue samples (see Table S1 and Table S8-9). Following immunostaining, high resolution images were acquired with a Zeiss LSM 710

Confocal Microscope. Collected images were imported into Imaris software Ver 7.6 (Bitplane, CH). Quantification of cell phenotype was performed by sampling five random fields (20X) taken from disease regions and the matched normal region in three patients. Firstly, 100-200 airway epithelial cells were counted based on DAPI/EpCAM staining. Cells that were SCGB1A1^{+} were also counted. From the same patients, the percentage of EpCAM^{+}proSPC^{+} cells relative to EpCAM^{+} cells were assessed from five random fields in each slide (20X).

2D air-liquid interface

Passage 3 EpCAM^{+}CD73^{+} epithelial cells were counted and single cells were seeded onto transwell inserts (0.4 µm Transwell insert, Corning) in expansion media on both the apical and basal side of the insert. After two days, media was removed from both the apical and basal surface and cells were grown at air-liquid interface (ALI) in airway differentiation media (PneumoCult™-ALI medium, StemCell Technologies). Regular media changes from the basal surface were made every three days. After 21 days, cells were fixed with 70% ethanol (Sigma). For immunofluorescence (IF), inserts were incubated with cooled solution of 95% ethanol / 5% glacial acetic acid for 10 minutes for permeabilization and washed 3X in TBS solution. Afterwards, inserts were blocked with 3% goat serum (blocking solution) for 1 hour at room temperature. Inserts were incubated overnight at 4°C with primary antibodies to detect mucous secreting cells, goblet, club and basal stem cells (see Table S6). Secondary antibodies (see Table S9) were diluted in washing solution and added to the inserts, which were incubated for 2 hours at room temperature. Nuclei were counterstained with DAPI. High resolution Z-axis images were acquired with a Zeiss LSM 710 Confocal Microscope. Images were

collected and imported into Imaris software Ver 7.6 (Bitplane, CH) to recreate three-dimensional (3D) volume reconstruction of the data set to visualize cell surface areas and volume performed using the Surpass tool. To examine the role of the NOTCH signaling pathway, separate wells were treated with delta-like ligand 4 (DLL4, 10 ng/ml, Peprotech) or γ-secretase inhibitor DAPT (20 μM, Sigma) or vehicle (DMSO, Sigma) at every media change and samples were processed for IF, as described above. From images, 5 random fields were counted in three technical replicates at 40X magnification. The total cell amounts based on the E-cadherin staining, then counted $SCGB1A1^+$, $MUC5AC^+$ and β-$tubulin^+$ cells respectively.

3D organoid culture

To generate airway organoids, single $EpCAM^+CD73^+$ cells from postnatal or adult lung were mixed with autologous $CD90^+$ stromal cells at a 1:1 ratio and resuspended in 50% solution of growth factor reduced matrigel (Corning) and seeded into inserts (0.4 μm Transwell insert, Corning). After solidification of matrigel:cell solution, airway differentiation media (PneumoCult™) was added to the basal chamber. Fresh media changes were made every 2 days for 21 days. To generate alveolar organoids, a 40% solution of growth factor reduced matrigel (Corning) was seeded into 0.4 μm Transwell inserts (Corning) and allowed to solidify forming a base. Afterwards, a 5% matrigel solution containing $EpCAM^+CD73^+$ cells with autologous $CD90^+$ stromal cells at a 1:1 ratio was seeded onto the top of matrigel base. 800 μl of expansion media was added to the basal chamber exposing cells to ALI. Transwell inserts were maintained in a humidified 37°C oxygen (21%) incubator in 5% CO_2 for expansion/differentiation. After 24 hours, media was removed from the apical surface and not replaced to mimic ALI,

whereas a fresh media change to the lower chamber was made using distal airway media (see Table S12) and media was changed every two days. After 21 days, inserts were fixed with 70% ethanol and processed for IF or removed from matrigel using dispase (Corning). High resolution Z-axis images were acquired with a Zeiss LSM 710 Confocal Microscope. Images were collected as lsm files and imported into Imaris software (Ver7.6) to recreate 3D volume reconstruction of the data set to visualize cell surface areas, organoid volume and cell-cell communication using the Surpass tool.

Transmission electron microscopy

Airway organoids were recovered from the matrigel using dispase (Corning) and submerged with fixative consisting of 2.5% glutaraldehyde (Agar Scientific, Stansted, Essex, UK) in 0.15M HEPES (Fluka, Buchs, Switzerland) with an osmolarity of 709 mOsm and adjusted to a pH of 7.34. The organoids remained in the fixative at 4°C for at least 24h, before being further processed. All samples were then washed with 0.15 M HEPES three times for 5 min, post fixed with 1% OsO_4 (SPI Supplies, West Chester, USA) in 0.1 M Na-cacodylate-buffer (Merck, Darmstadt, Germany) at 4°C for 1 h, washed with 0.05 M maleat-NaOH buffer (Merck, Darmstadt, Germany) three times for 5 min, and then block stained in 0.5% uranyl acetate (Fluka, Buchs, Switzerland) in 0.05 M maleat-NaOH buffer at 4°C for 1 h. Thereafter, cells were washed in 0.05 M maleat-NaOH buffer three times for 5 min and dehydrated in 70, 80 and 96% ethanol (Alcosuisse, Switzerland) for 15 min each at room temperature. Subsequently, cells were immersed in 100% ethanol (Merck, Darmstadt, Germany) three times for 10 min, in acetone (Merck, Darmstadt, Germany) two times for 10 min, and finally in acetone-epon (1:1) overnight at room temperature. The next day, cells were embedded in epon (Fluka,

Buchs, Switzerland) and left to harden at 60°C for 5 days. Sections were produced with an ultramicrotome UC6 (Leica Microsystems, Vienna, Austria), first semi-thin sections (1 um) for light microscopy which were stained with a solution of 0.5% toluidine blue O (Merck, Darmstadt, Germany) and then ultrathin sections (70-80 nm) for electron microscopy. The sections, mounted on single slot copper grids, were stained with uranyl acetate and lead citrate with an ultrostainer (Leica Microsystems, Vienna, Austria). Sections were then examined with a transmission electron microscope (CM12, Philips, Eindhoven) equipped with a digital camera (Morada, Soft Imaging System, Münster, Germany) and image analysis software (iTEM).

RNA extraction and real time quantitative PCR

Total RNA was extracted from cells or organoids using RNeasy Mini Kit (Qiagen) to analyze gene expression using real time quantitative PCR (RT-qPCR). RT-qPCR was performed in triplicates with target-specific primers using TaqMan Gene Expression Assay (Applied Biosystems) on AB7500 FAST real-time PCR system (Applied Biosystems). Expression levels were normalized to 3 internal controls tested for expression stability across samples in each experiment using Expression Suite Software (Life Technologies). Relative expression was calculated by $2^{-\Delta\Delta CT}$ method. (see Table S6 for primer list).

PD-L1 and CD47 immunohistochemistry

Serial sections from formalin-fixed and paraffin-embedded human lung adenocarcinoma (LUAD, n = 27) and lung squamous cell carcinoma (LUSC, n = 31) cases were stained for programmed death ligand-1 (PD-L1) and CD47. Immunohistochemical staining was

performed using an automated immunostainer (Bond III, Leica Biosystems, Muttenz, Switzerland) using the following antibodies: anti-human PD-L1 (clone E1L3N, Cell Signaling Technology, Damvers, MA, USA) at a dilution of 1:400 and anti-human CD47 (clone B6H12, Santa Cruz, San Diego, USA) at a dilution of 1:20. Sections were incubated with primary antibodies at room temperature for 15 minutes, followed by incubation with the secondary antibody using the Bond Polymer Refine Kit with 3-3'-Diaminobenzidine-DAB as chromogen (Leica Biosystems), counterstained with hematoxylin and mounted in Aquatex (Merck, Darmstadt, Germany). Membranous CD47 expression was scored 0-3 by a trained pathologist (SB). Tumoral PD-L1 expression was scored by a trained pathologist (SB) according to current guidelines as the percentage of cells with membranous staining of any intensity (tumor proportion score) and grouped as follows: <1%; between 1 to <50%; and ≥50%, as previously described (22) (see Table S5 and S6). For double immunohistochemistry of postnatal lung tissue, rabbit PD-L1 antibody (clone E1L3N) was diluted 1:400, incubated for 30 min. Samples were incubated with Horseradish Peroxidase (HRP)-polymer for 15 min and visualized using DAB as brown chromogen (Bond polymer refine detection, Leica Biosystems) for 10 min. Following this, slides were incubated with mouse ERG antibody (Agilent, M73149) diluted at 1:50 for 30 min. Following this, slides were incubated with secondary antibody Alkaline phosphatase (AP)-polymer for 15 min, and visualized using fast red as red chromogen (Red polymer refine Detection, Leica Biosystems). Samples were counterstained with haematoxylin and mounted with Aquatex (Merck). All slides were scanned and photographed using Pannoramic 250 (3DHistech).

Statistical analysis

Data are expressed as mean ± SD. Comparisons between two groups were carried out using the parametric student's two-tailed paired or unpaired t-test for normally distributed data. If data were not distributed normally, a nonparametric Wilcoxon signed-rank test was used between the two groups. All tests were two-tailed. Analysis of more than two groups was performed with ANOVA followed by Newman-Keuls post hoc test. The numbers of samples (biological replicates) per group (n), or the numbers of experiments (technical replicates) are specified in the figure legends. Data was analyzed using GraphPad Prism 8 software. Statistical significance is accepted at $P < 0.05$.

Results

$CD73^+$ labels a rare population of $EpCAM^+$ progenitor cells in both airway and alveolar region of unaffected human lung

As shown in the schematic panel in Figure 1A, we applied a multiparametric approach to identify and characterize resident lung epithelial progenitor cells in a cohort of pediatric patients (herein called postnatal) undergoing elective surgery for congenital lung lesions and other airway abnormalities (see Table S1) and adult patients diagnosed with NSCLC undergoing elective surgery for curative intent (see Table S3). We previously reported the presence of an $EpCAM^{neg}$ (gate R4) mesenchymal cluster further fractionated on the basis of CD73/CD90 in unaffected postnatal lung using polychromatic flow cytometry (49). Here, unlike cells in the $EpCAM^{neg}$, we show that the $EpCAM^{pos}$ fraction (gate R5, Figure 1B and Fig. S1A) contains a cluster of cells enriched for 5`ecto-nucleotidase CD73 (gate R6, 11.8±9.9%, Figure 1C and Supplemental Fig. S1A, D). In contrast, $CD73^+$ cells co-expressing membrane glycoprotein CD90 (THY-1) or single $CD90^+$ cells were rare ($CD73^+CD90^+$, 0.8±0.8% and $CD73^-CD90^+$, 3.9±2.8%, Figure 1D and Supplemental Fig. S1B-E). Surprisingly, there was no difference in the percent of $EpCAM^+CD73^+$ cells between postnatal and adult lung tissue (Figure 1E and Supplemental Fig. S1F).

Next, we wanted to localize $EpCAM^+CD73^+$ cells in postnatal lung. Apart from standard histological analysis based on H&E performed on unaffected lung tissue (Figure 1F-H), we immunostained separate sections with EpCAM and CD73. Confocal analysis revealed dual EpCAM/CD73 labelled cells occupying a basal position in the airway (Figure 1G, white arrow). In the alveolar region, cuboidal shaped cells co-staining for EpCAM and CD73 (white arrows) were found next to flat squamous-like $CD73^+$ cells

311 (white arrowhead) lacking EpCAM (Figure 1I). Further, EpCAM$^+$ cells in the airway co-
312 stain with the transcription factor SRY-Box 2 (SOX2), whereas in the alveolar region
313 were shown to express the ATII marker prosurfactant protein C (proSPC) (Supplemental
314 Fig. S1G-H).
315
316 **Prospectively isolated EpCAM$^+$CD73$^+$ cells resemble basal-like stem cells in**
317 **culture**
318 Next, using FACS we prospectively isolated EpCAM$^+$CD73$^+$ cells from unaffected
319 postnatal lung tissue and expanded cells in feeder-free plates using a chemically-
320 defined growth media (Figure 2A). At the mRNA level, EpCAM$^+$CD73$^+$ cells were
321 enriched for SOX2 and Keratin 5 (KRT5), as well as the basal stem cell transcription
322 factor p63 (TRP63) (Figure 2B). Expression of genes defining mature cell types were not
323 observed or low. Interestingly, there was a 2-fold increase in hypoxia inducible factor
324 HIF1α expression in EpCAM$^+$CD73$^+$ cells (Figure 2B). Immunostaining sorted cells after
325 reaching confluence in culture revealed that the majority of expanded EpCAM$^+$CD73$^+$
326 cells express the laminin receptor integrin β4, which supports cell adhesion between
327 basal epithelial cells and basement membrane (28). This was confirmed using flow
328 cytometry (Supplemental Fig. S2A). Variable expression for SOX2 and KRT5 were
329 observed, as well as lack of expression of proSPC and the club cell marker SCGB1A1
330 (Figure 2C-E). FACS analysis on independent postnatal specimens revealed that a
331 small fraction of postnatal EpCAM$^+$CD73$^+$ cells co-express integrin β4 *ex vivo*
332 (Supplemental Fig. S2B), which did not differ in adult lung (postnatal, 10.7±11% versus
333 adult, 7.1±5.6%, Supplemental Fig. S2C). Immunostaining of postnatal human lung
334 revealed integrin β4$^+$ cells located along the basal membrane of the conducting airway

co-expressing KRT5 (Supplemental Fig. S2D), whereas integrin $\beta4^+$ cells co-expressing proSPC in the alveolar region were rare (Supplemental Fig. S2E-F). To further immunophenotype the $EpCAM^+CD73^+$ population *in vivo*, we stained an additional six independent postnatal and seven adult tissue samples and show that a small subset within the $EpCAM^+CD73^+$ population co-express the ATII marker HTII-280 (15) in both postnatal (3.6±2.5%) and adult human lung (3.4±4.2%) (Figure 2F-G and Supplemental Fig. S2G, H). Moreover, sequential gating show variable expression of CD24 and PDPN within $EpCAM^+CD73^+HTII\text{-}280^-$ and $EpCAM^+CD73^+HTII\text{-}280^+$ subsets (Supplemental Fig. S2I-K). Taken together, $EpCAM^+CD73^+$ cells represent a rare but heterogeneous population in human lung. Our submersion culture conditions appear to favour the expansion of cells toward a $SOX2^+$ basal stem-cell like state (Figure 2H), consistent with primary sorted $Sox2^+EpCAM^+\beta4^+Krt5^-$ progenitor cells derived from murine lungs (50).

Generating airway epithelium from single $EpCAM^+CD73^+$ cells

Based on the location of $EpCAM^+CD73^+$ cells in the airway of postnatal lung and gene expression pattern following expansion in culture, we next investigated the airway differentiation capability of $EpCAM^+CD73^+$ cells from postnatal and adult lung tissue. Standard H&E of a lung section from the unaffected postnatal lung show normal lung structure including a bronchiole and surrounding alveolus (Figure 3A and Supplemental Fig. S3A). Using confocal microscopy, a single bronchiole is highlighted showing the airway epithelium stains for EpCAM/SOX2, whereas $KRT5^+$ cells (green) can be seen occupying the basal layer only (Figure 3B and Supplemental Fig. S3B). In separate sections, epithelial cells lining the basal membrane of the airway express the basal stem cell marker TRP63 (white arrowhead), some of which co-express KRT5 (white arrow,

Figure 3C and Supplemental Fig. S3C). We seeded single FACS-sorted EpCAM^{+}CD73^{+} cells at air-liquid-interface (ALI) to induce airway differentiation (Figure 3D). After 21 days, postnatal EpCAM^{+}CD73^{+} cells give rise to pseudostratified multiciliated-secretory epithelium (left panels, Figure 3D and Supplemental video S1). Persistent pharmacological activation of NOTCH signaling using the precanonical NOTCH ligand DLL4 during airway differentiation reduced both goblet (MUC5AC, white) and secretory club cell (SCGB1A1, purple) formation whereas cilia (β-tubulin, yellow) formation was intact (middle panels, Figure 3D and Supplemental Fig. S3D). Immunostaining with E-cadherin (green) and 3D volume rendering through the zy plane show an intact pseudostratified barrier and TRP63^{+} (red) cells along the basal membrane (Figure 3D-E). Pharmacological inhibition of NOTCH signaling using gamma secretase inhibitor DAPT also decreased the formation of both goblet cells and secretory club cells leaving cilia formation intact (far right panels, Figure 3D and Supplemental Fig. S3D). However, formation of a pseudostratified barrier was disrupted (Figure 3E). In comparison with adult-derived EpCAM^{+}CD73^{+} cells, we noted several differences (Figure 3F). First, adult-derived EpCAM^{+}CD73^{+} cells required the presence of autologous CD90^{+} stromal cells to ensure proper airway differentiation (left panels, Figure 3F and Supplemental video S2). Second, the formation of a mucociliary epithelium was intact despite persistent NOTCH signaling (middle panels Figure 3F and Supplemental Fig. S3E). Third, inhibition of NOTCH signaling reduced cilia formation (Figure 3F and Supplemental Fig. S3E), yet the disruption in the formation of pseudostratified barrier was similar with postnatal lung EpCAM^{+}CD73^{+} cells (Figure 3G).

EpCAM$^+$CD73$^+$ cells generate three-dimensional organoid structures with airway and alveolar-like features

The self-organizing feature possessed by adult stem cells is indispensable for the formation of organoids, which are three dimensional (3D) structures recapitulating important characteristics of the organ they represent (32). Thus, we next asked whether EpCAM$^+$CD73$^+$ cells could give rise to organoids that recapitulate human airway found *in vivo*. We seeded single EpCAM$^+$CD73$^+$ cells together with single autologous CD90$^+$ mesenchymal cells from the postnatal or adult lung in 3D matrigel, exposing the apical surface to air while culturing in human airway differentiation media in the basal chamber (Figure 4A). After 21 days, transmission electron microscopy demonstrated that organoid structures were morphologically similar to *in vivo* pseudostratified mucociliary epithelium (Figure 4B). Ciliated cells and secretory cells can be found facing inside the lumen of the organoid structure and small basal-like cells in the basal layer. At the mRNA level, airway organoid structures upregulated genes found to be enriched in airway (Figure 4C). Organoids with beating cilia can be detected after removal from matrigel (Supplemental video 3 and 4).

Next, we explored whether EpCAM$^+$CD73$^+$ cells could generate organoids with alveolar-like features. Single postnatal EpCAM$^+$CD73$^+$ cells with autologous CD90$^+$ stromal cells were suspended in a solution of matrigel (5%), which was overlaid on a base of matrigel (40%) and grown in a modified alveolar induction media placed in the basal chamber biased towards alveolar differentiation (Figure 4D). After 21 days, confocal imaging of an organoid shows the presence of rare ATII-like cells positive for proSPC at the apical surface of the lumen (white arrow, Figure 4E and Supplemental Fig. S4A-B). In separate organoids, rare cells positive for the ATI marker HOPX can be

observed, which were distinct from KRT5^{+} stained cells (Figure 4F and Supplemental Fig. S4C). Similar results were observed for alveolar organoids derived from adult EpCAM^{+}CD73^{+} cells (Supplementary Fig. S4D-E). FACS analysis revealed the presence of EpCAM^{+}CD73^{+}HTII-280^{-} and EpCAM^{+}CD73^{+}HTII-280^{+} cells using FACS at levels similar to those found *in vivo* (Figure 4G, H). However, the level of single positive HTII-280 epithelial cells was diminished (organoids, 13.42±4.84% and postnatal lung, 55.3±26.7%). Co-expression of CD24 and PDPN were upregulated in alveolar organoids (Figure 4I, J). Moreover, a greater percent of EpCAM^{+}CD73^{+} cells co-expressed β4^{+} cells compared with *in vivo* native postnatal lung (organoids, 40.6±31% and postnatal lung, 11.55±15%) (Supplemental Fig. S4F-G).

Expansion of EpCAM^{+}CD73^{+} cells together with mature epithelial cell lineages in congenital lung lesions and other airway abnormalities

To determine whether the EpCAM^{+}CD73^{+} cell subset or other cell lineages are induced after lung injury, we performed immunofluorescence analysis of matched lung tissue obtained from patients diagnosed with congenital lung lesions and other airway abnormalities (Table S1). Standard H&E staining show structural airway malformations and cystic lesions from representative patients (Figure 5A-D and Supplemental Fig. S5). In highlighted regions of H&E sections (Figure 5E, G), clusters of EpCAM^{+}CD73^{+} cells in the airway of a patient with CPAM (Figure 5F) or chronic bronchiolitis (Figure 5H) can be observed with confocal imaging. In a separate patient with congenital intrapulmonary sequestration, H&E shows the dysplastic alveolar epithelium (Figure 5I) and corresponding confocal image shows the extent of EpCAM/CD73 localization (Figure 5J). There was no evidence of clusters of TRP63^{+}KRT5^{+} cells migrating from the

proximal airway or ectopic TRP63^{+}KRT5^{+} or KRT5^{+} cells in the distal lung between cystic lesions or in dysplastic alveolar regions in CPAM tissue (Figure S6D-F). Moreover, in these same patients, EpCAM^{+}SOX2^{+} or EpCAM^{+}SOX2^{+}KRT5^{+} cells were observed in the basal layer of cysts but not in the thickened interstitium (Supplemental Fig. S6G-I).

Confocal imaging analysis demonstrated that cystic lesions and areas of dysplastic alveolar epithelium were lined with hyperplastic EpCAM^{+}proSPC^{+} cells (yellow arrow), either as single cells or as clusters in patients with congenital lung lesions (Figure 6A-F). In other regions, bronchiolar lined cysts of cuboidal and columnar epithelium stained positive for EpCAM^{+}SCGB1A1^{+} (yellow arrowhead), as well as lining the airways in the damaged region of the lung (Figure 6A-F). We noted a significant increase in both EpCAM^{+}proSPC^{+} and EpCAM^{+}SCGB1A1^{+} cell subtypes (Figure 6G). In cystic lung lesions we did not observe dual positive proSPC-SCGB1A1 cells indicative of putative bronchioalveolar stem/progenitor cell with regenerative potential, as described in various mouse models of lung injury (29, 40).

EpCAM^{+}CD73^{+} cells re-emerge in lung adenocarcinoma and upregulate immune checkpoint molecules PD-L1 and CD47

CD73 has emerged as a novel therapeutic target in solid tumors due its role in the enzymatic generation of the immunosuppressive molecule adenosine (1). We investigated whether there is re-emergence of CD73 expression on tumor epithelial cells marked by EpCAM in NSCLC and whether this cell cluster expresses additional inhibitory immune checkpoint molecules involved in immune resistance. To accomplish this, we profiled the composition of tumor epithelium (EpCAM^{+}) and matched uninvolved

lung tissue using polychromatic flow cytometry in a cohort of 122 surgically resected stage I to IV NSCLC tissues. Clinical and pathologic characteristics for all patients are found in Table S3. In representative patients, H&E staining shows tumor islands surrounded by stroma (Figure 7A, C and Supplemental Fig. S7). Confocal imaging demonstrated that $EpCAM^+$ tumor islands co-stain for CD73 (Figure 7B). A serial section confirmed that these cells also were TRP63 positive (Figure 7D and Supplemental Fig. S7B-C, E). Flow cytometric analysis demonstrated an enrichment in $EpCAM^+$ tumor cells co-expressing CD73 in lung adenocarcinoma (LUAD) (Figure 7E). In the lung squamous cell carcinoma (LUSC) cohort, there was an enrichment in $EpCAM^+CD73^+$ cells in tumor in subsets of patients; however, overall this was not significant (Figure 7F). The inhibitory immune checkpoint molecule PD-L1 was upregulated in the $EpCAM^+CD73^+$ TC fraction in both LUAD and LUSC (Figure 7G, H). Interestingly, membrane expression of CD47, a “don’t eat me signal” and inhibitory checkpoint molecule of the innate immune system, was found to be overexpressed in the $EpCAM^+CD73^+$ TC fraction only in LUAD (Figure 7G, H). Comparison of LUAD with LUSC revealed a difference only in $EpCAM^+CD73^+$ cells (Supplemental Fig. S7F-I). To determine the level of colocalization of tumoral PD-L1 and CD47, we performed immunohistochemistry on serial sections of 58 NSCLC cases (LUAD, n = 27; LUSC, n = 31). Representative micrographs of serial sections capture the heterogeneity in tumoral PD-L1 and CD47 co-expression (Figure 7I), which was depicted in an expanded number of cases (Supplemental Fig. S7J-K) and quantified in table format (Supplemental Table S5 and S6). Surprisingly, we did not detect PD-L1 expression within the epithelium of congenital lung lesions we examined despite immune infiltration being present in some case (Supplemental Fig. S8).

Based on the compositional features identified in LUAD and LUSC the following 6 markers CD73 (NT5E), EpCAM (EPCAM), CD90 (THY1), PD-L1 (CD274), CD47 and CD127 (IL-7R) were used to determine their clinical relevance in predicting patient survival using mRNA expression data from The Cancer Gene Atlas (TCGA) NSCLC cohort. We were able to build a compound Cox proportional-hazards model to find the best combination of the 6-gene signature associated with survival. In LUAD, we found a compound model whereby increased expression of CD73 at the mRNA level was associated with shorter patient survival, whereas this was not the case for LUSC (Figure 7J). Taken together, these data highlight the importance of CD73 as a potential novel drug target in adenocarcinoma histology.

Discussion

Here, we show that CD73 enriches for a rare EpCAM$^+$ cell subset isolated from healthy postnatal and adult human lung tissue that can give rise to pseudostratified mucociliary epithelium, and to a limited extent, mature alveolar cell types. Spatially, EpCAM/CD73 double positive cells were found in the respiratory epithelium and alveolar region. Importantly, this rare population remains stable during lung maturation. In congenital cystic lung lesions and other airway abnormalities foci of EpCAM$^+$CD73$^+$ cells line the airway epithelium. Interestingly, in LUAD EpCAM$^+$CD73$^+$ cells re-emerge and co-express CD47 and PD-L1, proteins that negatively regulate host innate and adaptive immune responses, respectively, and are known to contribute to tumor immune escape.

A significant challenge in the field is the use of surface-marker based phenotyping to identify cell types in human lung endowed with stem cell function under stable conditions and during perturbations. Using EpCAM as a surrogate marker of normal lung stem cells, Hogan and colleagues were the first to demonstrate that EpCAM$^+$ cells co-expressing the ATII maker HTII-280 (15) function as bona fide ATII stem cells when cocultured with niche cells (2). Following this, numerous studies have since confirmed the use of EpCAM to isolate lung progenitors from human iPSCs to generate 3D lung organoids with both airway and alveolar features (8, 12, 16), and to isolate ATII cells for downstream characterization within adult human lung (7, 9) Using a flow cytometric approach, we demonstrate that a small subset of postnatal lung EpCAM$^+$CD73$^+$ cells co-express this HTII-280, which does not change with age. The alveolar differentiation potential of EpCAM$^+$CD73$^+$ cells was rare, despite immunostaining data showing anatomically distinct cuboidal shaped EpCAM$^+$CD73$^+$ cells in the alveolar region. In sharp contrast, 2D or 3D airway differentiation potential of culture expanded

512 EpCAM$^+$CD73$^+$ cells was robust. A limitation may be the inability to expand bona fide
513 human alveolar progenitor cells, which are SFPC$^+$ (18) and the intervening period of
514 submersion culture *in vitro*. In murines, submersion culture induces hypoxia-driven
515 hyperactive NOTCH signaling that favour the expansion of cells toward a Sox2$^+$ basal
516 stem-cell like state and impairs their alveolar differentiation potential (50). We observed
517 that 11% of EpCAM$^+$CD73$^+$ cells *in vivo* express integrin β4, yet after submersion culture
518 integrin β4 was universally expressed. Therefore, 2D *in vitro* culture may select for
519 airway-derived EpCAM$^+$CD73$^+$ cells that are committed towards a basal cell state.
520 Interestingly, a fraction of EpCAM$^+$CD73$^+$HTII-280$^+$ cells were recovered from alveolar
521 organoids similar to what was observed in native lung. Despite this, we noted that the
522 percent of EpCAM$^+$HTII-280$^+$ cells was considerably lower compared with native lung.
523 Alveolar organoids were generated with autologous CD90$^+$ mesenchymal cells in a
524 media that has been shown to favour alveospheres (2). Whether CD90$^+$ mesenchymal
525 cells this impairs alveolar differentiation due to excessive TGF-β1 activation originating
526 from the mesenchyme compartment requires further investigation (35).

527 CPAM includes a wide range of developmental lung malformations arising *in*
528 *utero* marked by cystic and/or adenomatous pulmonary areas, which was consistent with
529 previous reports (14, 26, 43, 44). In between cystic airspaces, the distal lung is marked
530 by thickened interstitial spaces lined by simple epithelium with expanding EpCAM$^+$ cells
531 expressing the club cell marker secretoglobin SCGB1A1, whereas dysplastic alveolar
532 epithelium is marked by cuboidal EpCAM$^+$ cells coexpressing proSPC. We did not
533 observe SCGB1A1$^+$proSPC$^+$ putative bronchioalveolar stem cells nor regenerating pods
534 of TRP63$^+$KRT5$^+$ cells, as has been described in murine models of lung injury and
535 cancer (23, 24, 30, 39, 40, 48). Although incompletely understood, CPAM is thought to

arise due to altered branching morphogenesis during fetal lung development. Recent evidence based on transcriptome-wide analysis of congenital lung lesions revealed dysregulated expression of genes related to RAS and PI3K-AKT-mTOR pathway together with a cell-autonomous defect in growth and airway differentiation of isolated $EpCAM^+$ cells (44). (49). EpCAM is not restricted to the respiratory tract, however and whether these cells were enriched in CD73 and the mechanism driving their expansion was not described. In human fetal lung explants, coordination between highly proliferative dual positive SOX2-SOX9 progenitor cells located at distal branching tips with smooth muscle cells (SMC) in time and space is implicated in proper branching morphogenesis (10). Importantly, we show that $EpCAM^+CD73^+$ cells co-express both SOX2 and SOX9 at the mRNA level. We previously demonstrated in CPAM that thick interstitial spaces were filled with mesenchymal cells (49). It is presently unclear whether the mesenchyme plays a causative role in congenital lung lesions and requires further investigation.

The natural evolution of CPAM also posses an increased risk for malignancy, although the true incidence is still not known (26). Lung tumors associated with CPAM in children range from rhabdomyosarcoma (RMS), pleuro-pulmonary blastoma (PPB), whereas in the adult, bronchioalveolar carcinoma (BAC) and adenocarcinoma are more common (6). Goblet cell proliferation, which has been described in CPAM may represent a precursor lesion to lung adenocarcinoma in children (13). Along these lines, various changes in several notable genes including FGF10, FGFR2b, SOX2 and mutations in KRAS at codon 12 was associated with adenocarcinoma. Recently, whole exome sequencing a cohort of eighteen CPAM patients revealed mutations linked to lung development and cancer development (19). As eluded to above, one of the main

560 difficulties has been in inability to identify the cell of origin of lung cancer in humans.
561 Moving forward it will be necessary to perform single-cell transcriptome analysis and
562 differentiation trajectory inference to uncover the fate of $EpCAM^{+}CD73^{+}$ cells when
563 moving from normal to aged and CPAM lung (51).

564 Regulatory networks important during development can re-emerge after tissue
565 injury and malignant transformation (17). One such regulatory molecule is CD73. Our
566 data in LUAD along with findings from breast (4) and ovarian cancer (47) suggest that
567 CD73 represents a critical target in solid tumors. Functionally, this cell type ectoenzyme
568 CD73 involved in generation of the signaling molecule adenosine via dephosphorylation
569 of adenosine monophosphate. Production of extracellular adenosine functions as an
570 immune suppressor initiating a cascade of events that counterbalance pro-inflammation.
571 Under chronic inflammatory conditions, adenosine signalling can be maladaptive
572 contributing to tumor immune escape (42). Although we did not correlate the increased
573 tumoral CD73 expression with tumor genotype, Nakagawa and colleagues reported
574 CD73 expression on tumor cells in *EGFR*-mutation positive NSCLC increased after
575 targeted treatment in patients with previously high PD-L1 expression (21). In triple
576 negative breast cancer, chemotherapy enriches for a subset of tumor cells co-
577 expressing CD73/CD47/PD-L1 with immune evasive properties regulated, in part, via
578 HIF1α (41). A limitation in our study is that we did not correlate increased membrane
579 expression with enzymatic activity and extracellular adenosine production. In NSCLC,
580 immune checkpoint inhibitors targeting PD-L1/PD-1 axis have shown clinically
581 meaningful response rates in approximately 20% of patients; however, the majority of
582 patients still derive no therapeutic benefit from immune checkpoint blockade (11).
583 Recently, dual targeting of CD47 and PD-L1 on tumor cells in immunocompetent

584 preclinical mouse models showed enhanced therapeutic efficacy in controlling tumor
585 growth, in part, via re-invigoration of the host immune system (27, 31). Therefore,
586 targeting CD73 in combination with immune checkpoint blockers targeting PD-L1 and
587 CD47 might represent a novel treatment strategy in a subset of patients with LUAD (20,
588 36) that requires further investigation.
589

ACKNOWLEDGEMENTS

Tissue were provided by the Tissue Bank Bern. Electron microscopy sample preparation and imaging were performed with devices supported by the Microscopy Imaging Center (MIC) of the University of Bern. We thank Barbara Krieger from the Institute of Anatomy, University of Bern for preparation of the EM figures. We thank staff from the FACSLab Core facility, Department of BioMedical Research, University of Bern for their assistance in performing the sorting experiments.

AUTHOR CONTRIBUTIONS: conception & design – LW, SRRH; Data acquisition – LW, PD, CS, LF, SB, STS, BH, FB, CW, RWP, TMM, PKB, UM, RAS, SRRH; Data interpretation & analysis – LW, PD, SB, STS, PKB, UM, SRRH; Drafting of Manuscript – LW, SRRH; Editing of manuscript - Final Approval of manuscript – LW, RAS, SRRH.

References

1. **Antonioli L, Yegutkin GG, Pacher P, Blandizzi C, and Hasko G**. Anti-CD73 in cancer immunotherapy: awakening new opportunities. *Trends in cancer* 2: 95-109, 2016.
2. **Barkauskas CE, Cronce MJ, Rackley CR, Bowie EJ, Keene DR, Stripp BR, Randell SH, Noble PW, and Hogan BL**. Type 2 alveolar cells are stem cells in adult lung. *The Journal of clinical investigation* 123: 3025-3036, 2013.
3. **Borea PA, Gessi S, Merighi S, Vincenzi F, and Varani K**. Pathological overproduction: the bad side of adenosine. *British journal of pharmacology* 174: 1945-1960, 2017.
4. **Buisseret L, Pommey S, Allard B, Garaud S, Bergeron M, Cousineau I, Ameye L, Bareche Y, Paesmans M, Crown JPA, Di Leo A, Loi S, Piccart-Gebhart M, Willard-Gallo K, Sotiriou C, and Stagg J**. Clinical significance of CD73 in triple-negative breast cancer: multiplex analysis of a phase III clinical trial. *Annals of oncology : official journal of the European Society for Medical Oncology* 29: 1056-1062, 2018.
5. **Butler JP, Loring SH, Patz S, Tsuda A, Yablonskiy DA, and Mentzer SJ**. Evidence for adult lung growth in humans. *The New England journal of medicine* 367: 244-247, 2012.
6. **Casagrande A, and Pederiva F**. Association between Congenital Lung Malformations and Lung Tumors in Children and Adults: A Systematic Review. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 11: 1837-1845, 2016.
7. **Castaldi A, Horie M, Rieger ME, Dubourd M, Sunohara M, Pandit K, Zhou B, Offringa IA, Marconett CN, and Borok Z**. Genome-wide integration of microRNA and transcriptomic profiles of differentiating human alveolar epithelial cells. *American journal of physiology Lung cellular and molecular physiology* 2020.
8. **Chen YW, Huang SX, de Carvalho A, Ho SH, Islam MN, Volpi S, Notarangelo LD, Ciancanelli M, Casanova JL, Bhattacharya J, Liang AF, Palermo LM, Porotto M, Moscona A, and Snoeck HW**. A three-dimensional model of human lung development and disease from pluripotent stem cells. *Nature cell biology* 19: 542-549, 2017.
9. **Correll KA, Edeen KE, Zemans RL, Redente EF, Serban KA, Curran-Everett D, Edelman BL, Mikels-Vigdal A, and Mason RJ**. Transitional human alveolar type II epithelial cells suppress extracellular matrix and growth factor gene expression in lung fibroblasts. *American journal of physiology Lung cellular and molecular physiology* 2019.
10. **Danopoulos S, Alonso I, Thornton ME, Grubbs BH, Bellusci S, Warburton D, and Al Alam D**. Human lung branching morphogenesis is orchestrated by the spatiotemporal distribution of ACTA2, SOX2, and SOX9. *American journal of physiology Lung cellular and molecular physiology* 314: L144-L149, 2018.
11. **Doroshow DB, Sanmamed MF, Hastings K, Politi K, Rimm DL, Chen L, Melero I, Schalper KA, and Herbst RS**. Immunotherapy in Non-Small Cell Lung Cancer: Facts and Hopes. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2019.
12. **Dye BR, Hill DR, Ferguson MA, Tsai YH, Nagy MS, Dyal R, Wells JM, Mayhew CN, Nattiv R, Klein OD, White ES, Deutsch GH, and Spence JR**. In vitro generation of human pluripotent stem cell derived lung organoids. *eLife* 4: 2015.

13. **Fakler F, Aykutlu U, Brcic L, Eidenhammer S, Thueringer A, Kashofer K, Kulka J, Timens W, and Popper H**. Atypical goblet cell hyperplasia occurs in CPAM 1, 2, and 3, and is a probable precursor lesion for childhood adenocarcinoma. *Virchows Archiv : an international journal of pathology* 2019.
14. **Fowler DJ, and Gould SJ**. The pathology of congenital lung lesions. *Seminars in pediatric surgery* 24: 176-182, 2015.
15. **Gonzalez RF, Allen L, Gonzales L, Ballard PL, and Dobbs LG**. HTII-280, a biomarker specific to the apical plasma membrane of human lung alveolar type II cells. *J Histochem Cytochem* 58: 891-901, 2010.
16. **Gotoh S, Ito I, Nagasaki T, Yamamoto Y, Konishi S, Korogi Y, Matsumoto H, Muro S, Hirai T, Funato M, Mae S, Toyoda T, Sato-Otsubo A, Ogawa S, Osafune K, and Mishima M**. Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. *Stem cell reports* 3: 394-403, 2014.
17. **Herlyn D, Herlyn M, Steplewski Z, and Koprowski H**. Monoclonal antibodies in cell-mediated cytotoxicity against human melanoma and colorectal carcinoma. *European journal of immunology* 9: 657-659, 1979.
18. **Hiemstra PS, Tetley TD, and Janes SM**. Airway and alveolar epithelial cells in culture. *The European respiratory journal* 54: 2019.
19. **Hsu JS, Zhang R, Yeung F, Tang CSM, Wong JKL, So MT, Xia H, Sham P, Tam PK, Li M, Wong KKY, and Garcia-Barcelo MM**. Cancer gene mutations in congenital pulmonary airway malformation patients. *ERJ open research* 5: 2019.
20. **Ishii H, Azuma K, Kinoshita T, Matsuo N, Naito Y, Tokito T, Yamada K, and Hoshino T**. Predictive value of CD73 expression in EGFR-mutation positive non-small-cell lung cancer patients received immune checkpoint inhibitors. *Journal of Clinical Oncology* 36: 9065-9065, 2018.
21. **Isomoto K, Haratani K, Hayashi H, Shimizu S, Tomida S, Niwa T, Yokoyama T, Fukuda Y, Chiba Y, Kato R, Tanizaki J, Tanaka K, Takeda M, Ogura T, Ishida T, Ito A, and Nakagawa K**. Impact of EGFR-TKI Treatment on the Tumor Immune Microenvironment in EGFR Mutation-Positive Non-Small Cell Lung Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2020.
22. **Keller MD, Neppl C, Irmak Y, Hall SR, Schmid RA, Langer R, and Berezowska S**. Adverse prognostic value of PD-L1 expression in primary resected pulmonary squamous cell carcinomas and paired mediastinal lymph node metastases. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 31: 101-110, 2018.
23. **Kim CF, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, Crowley D, Bronson RT, and Jacks T**. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* 121: 823-835, 2005.
24. **Kumar PA, Hu Y, Yamamoto Y, Hoe NB, Wei TS, Mu D, Sun Y, Joo LS, Dagher R, Zielonka EM, Wang de Y, Lim B, Chow VT, Crum CP, Xian W, and McKeon F**. Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. *Cell* 147: 525-538, 2011.
25. **Leach JP, and Morrisey EE**. Repairing the lungs one breath at a time: How dedicated or facultative are you? *Genes & development* 32: 1461-1471, 2018.

26. **Leblanc C, Baron M, Desselas E, Phan MH, Rybak A, Thouvenin G, Lauby C, and Irtan S**. Congenital pulmonary airway malformations: state-of-the-art review for pediatrician's use. *European journal of pediatrics* 176: 1559-1571, 2017.
27. **Lian S, Xie R, Ye Y, Xie X, Li S, Lu Y, Li B, Cheng Y, Katanaev VL, and Jia L**. Simultaneous blocking of CD47 and PD-L1 increases innate and adaptive cancer immune responses and cytokine release. *EBioMedicine* 42: 281-295, 2019.
28. **Litjens SH, de Pereda JM, and Sonnenberg A**. Current insights into the formation and breakdown of hemidesmosomes. *Trends in cell biology* 16: 376-383, 2006.
29. **Liu Q, Liu K, Cui G, Huang X, Yao S, Guo W, Qin Z, Li Y, Yang R, Pu W, Zhang L, He L, Zhao H, Yu W, Tang M, Tian X, Cai D, Nie Y, Hu S, Ren T, Qiao Z, Huang H, Zeng YA, Jing N, Peng G, Ji H, and Zhou B**. Author Correction: Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. *Nature genetics* 51: 766, 2019.
30. **Liu Q, Liu K, Cui G, Huang X, Yao S, Guo W, Qin Z, Li Y, Yang R, Pu W, Zhang L, He L, Zhao H, Yu W, Tang M, Tian X, Cai D, Nie Y, Hu S, Ren T, Qiao Z, Huang H, Zeng YA, Jing N, Peng G, Ji H, and Zhou B**. Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. *Nature genetics* 51: 728-738, 2019.
31. **Liu X, Liu L, Ren Z, Yang K, Xu H, Luan Y, Fu K, Guo J, Peng H, Zhu M, and Fu YX**. Dual Targeting of Innate and Adaptive Checkpoints on Tumor Cells Limits Immune Evasion. *Cell reports* 24: 2101-2111, 2018.
32. **Miller AJ, and Spence JR**. In Vitro Models to Study Human Lung Development, Disease and Homeostasis. *Physiology* 32: 246-260, 2017.
33. **Minor M, Alcedo KP, Battaglia RA, and Snider NT**. Cell type- and tissue-specific functions of ecto-5'-nucleotidase (CD73). *American journal of physiology Cell physiology* 317: C1079-C1092, 2019.
34. **Munz M, Baeuerle PA, and Gires O**. The emerging role of EpCAM in cancer and stem cell signaling. *Cancer research* 69: 5627-5629, 2009.
35. **Ng-Blichfeldt JP, de Jong T, Kortekaas RK, Wu X, Lindner M, Guryev V, Hiemstra PS, Stolk J, Konigshoff M, and Gosens R**. TGF-beta activation impairs fibroblast ability to support adult lung epithelial progenitor cell organoid formation. *American journal of physiology Lung cellular and molecular physiology* 317: L14-L28, 2019.
36. **Park LC, Rhee K, Kim WB, Cho A, Song J, Anker JF, Oh M, Bais P, Namburi S, Chuang J, and Chae YK**. Immunologic and clinical implications of CD73 expression in non-small cell lung cancer (NSCLC). *Journal of Clinical Oncology* 36: 12050-12050, 2018.
37. **Picher M, Burch LH, Hirsh AJ, Spychala J, and Boucher RC**. Ecto 5'-nucleotidase and nonspecific alkaline phosphatase. Two AMP-hydrolyzing ectoenzymes with distinct roles in human airways. *The Journal of biological chemistry* 278: 13468-13479, 2003.
38. **Prasad M, Kumar B, Bhat-Nakshatri P, Anjanappa M, Sandusky G, Miller KD, Storniolo AM, and Nakshatri H**. Dual TGFbeta/BMP Pathway Inhibition Enables Expansion and Characterization of Multiple Epithelial Cell Types of the Normal and Cancerous Breast. *Molecular cancer research : MCR* 17: 1556-1570, 2019.

39. **Ray S, Chiba N, Yao C, Guan X, McConnell AM, Brockway B, Que L, McQualter JL, and Stripp BR**. Rare SOX2+ Airway Progenitor Cells Generate KRT5+ Cells that Repopulate Damaged Alveolar Parenchyma following Influenza Virus Infection. *Stem cell reports* 7: 817-825, 2016.
40. **Salwig I, Spitznagel B, Vazquez-Armendariz AI, Khalooghi K, Guenther S, Herold S, Szibor M, and Braun T**. Bronchioalveolar stem cells are a main source for regeneration of distal lung epithelia in vivo. *The EMBO journal* 38: 2019.
41. **Samanta D, Park Y, Ni X, Li H, Zahnow CA, Gabrielson E, Pan F, and Semenza GL**. Chemotherapy induces enrichment of CD47(+)/CD73(+)/PDL1(+) immune evasive triple-negative breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* 115: E1239-E1248, 2018.
42. **Sidders B, Zhang P, Goodwin K, O'Connor G, Russell DL, Borodovsky A, Armenia J, McEwen R, Linghu B, Bendell JC, Bauer TM, Patel MR, Falchook GS, Merchant M, Pouliot G, Barrett JC, Dry JR, Woessner R, and Sachsenmeier K**. Adenosine Signaling Is Prognostic for Cancer Outcome and Has Predictive Utility for Immunotherapeutic Response. *Clinical cancer research : an official journal of the American Association for Cancer Research* 26: 2176-2187, 2020.
43. **Stocker JT, Madewell JE, and Drake RM**. Congenital cystic adenomatoid malformation of the lung. Classification and morphologic spectrum. *Human pathology* 8: 155-171, 1977.
44. **Swarr DT, Peranteau WH, Pogoriler J, Frank DB, Adzick NS, Hedrick HL, Morley M, Zhou S, and Morrisey EE**. Novel Molecular and Phenotypic Insights into Congenital Lung Malformations. *American journal of respiratory and critical care medicine* 2018.
45. **Toufen C, Jr., Costa EL, Hirota AS, Li HY, Amato MB, and Carvalho CR**. Follow-up after acute respiratory distress syndrome caused by influenza a (H1N1) virus infection. *Clinics (Sao Paulo)* 66: 933-937, 2011.
46. **Trzpis M, McLaughlin PM, de Leij LM, and Harmsen MC**. Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. *The American journal of pathology* 171: 386-395, 2007.
47. **Turcotte M, Spring K, Pommey S, Chouinard G, Cousineau I, George J, Chen GM, Gendoo DM, Haibe-Kains B, Karn T, Rahimi K, Le Page C, Provencher D, Mes-Masson AM, and Stagg J**. CD73 is associated with poor prognosis in high-grade serous ovarian cancer. *Cancer research* 75: 4494-4503, 2015.
48. **Vaughan AE, Brumwell AN, Xi Y, Gotts JE, Brownfield DG, Treutlein B, Tan K, Tan V, Liu FC, Looney MR, Matthay MA, Rock JR, and Chapman HA**. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature* 517: 621-625, 2015.
49. **Wang L, Dorn P, Zeinali S, Froment L, Berezowska S, Kocher GJ, Alves MP, Brugger M, Esteves BIO, Blank F, Wotzkow C, Steiner S, Amacker M, Peng RW, Marti TM, Guenat OT, Bode PK, Moehrlen U, Schmid RA, and Hall SRR**. CD90(+)CD146(+) identifies a pulmonary mesenchymal cell subtype with both immune modulatory and perivascular-like function in postnatal human lung. *American journal of physiology Lung cellular and molecular physiology* 318: L813-L830, 2020.
50. **Xi Y, Kim T, Brumwell AN, Driver IH, Wei Y, Tan V, Jackson JR, Xu J, Lee DK, Gotts JE, Matthay MA, Shannon JM, Chapman HA, and Vaughan AE**. Local

lung hypoxia determines epithelial fate decisions during alveolar regeneration. *Nature cell biology* 19: 904-914, 2017.
51. **Zaragosi LE, Deprez M, and Barbry P**. Using single-cell RNA sequencing to unravel cell lineage relationships in the respiratory tract. *Biochemical Society transactions* 48: 327-336, 2020.
52. **Zunder ER, Lujan E, Goltsev Y, Wernig M, and Nolan GP**. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell stem cell* 16: 323-337, 2015.

Figure Legends

Figure 1. EpCAM$^+$ cells enriched for CD73 found in both respiratory epithelium and alveolar region of postnatal lung. (A) Illustration of approach to identify and characterize epithelial cell subsets in human lung. (B-C) Representative flow plots are shown. (D) Scatter plots show percentage of cell subsets within EpCAM$^+$ fraction (gate R4) after subgating for CD73 and CD90 in postnatal lung. (E) Scatter plots comparing EpCAM$^+$CD73$^+$ cell subset in postnatal versus adult human lung. N = 19, biological replicates for postnatal lung; N = 15, biological replicates for adult lung. (F, H) Hematoxylin and Eosin (H&E) stained unaffected postnatal lung. (G, I) Immunostaining of postnatal lung showing EpCAM (red) cells co-expressing CD73 (green, white arrow) in respiratory epithelium (G) and alveolar region (I). In the alveolus, CD73 cells lacking EpCAM are also found (white arrowhead). Nuclei were counterstained with DAPI. Br, bronchiole; Alv, alveolar; Lu, bronchiolar lumen. Scale bars 30 μm (G) and 20 μm (I). Data are presented as mean ± SD. Error bars show SD. *P* values are shown in the figure.

Figure 2. EpCAM$^+$CD73$^+$ cells are basal cell-like after expansion in culture. (A) Diagram of assay to expand prospectively isolated EpCAM$^+$CD73$^+$ cells. Phase contrast image of cells after reaching confluence. (B) mRNA levels of genes marking proximal and distal airway cells and markers of basal stem cells, alveolar type II and type I cells in FACS-sorted EpCAM$^+$CD73$^+$ cells by RT-qPCR. Upper airway (n = 3, biological replicates), postnatal lung tissue (n = 5, biological replicates) and EpCAM$^+$CD73$^+$ cells (n = 9, biological replicates). Gene expression level in postnatal lung tissue is set at one. Representative immunofluorescence stains of FACS-sorted EpCAM$^+$CD73$^+$ cells (C, D)

and their quantification (E) after reaching confluence in culture (n = 3, biological replicates). Scale bar 50 μm (C, D). (F) Representative FACS plots demonstrating that HTII-280 marks a minor subset of $EpCAM^{+}CD73^{+}$ progenitor cells (gate R6). (G) Scatter plots showing co-expression of CD24 and PDPN. (postnatal, n = 6, biological replicates; adult, n = 7, biological replicates). (H) Schematic depicting the different cell subsets found within the $EpCAM^{+}CD73^{+}$ fraction after expansion in culture. Data are presented as mean ± SD. Error bars show SD. *$P < 0.05$; ns, not significant; ND, not detected.

Figure 3. Disrupting epithelial NOTCH signaling alters mucociliary-secretory cell fate of $EpCAM^{+}CD73^{+}$ cells. (A) Representative hematoxylin and eosin (H&E) stain of unaffected lung showing normal bronchiole (Br) and surrounding alveolar region (Alv). (B) Immunostaining postnatal lung for TRP63 (red) and KRT5 (green). Higher-power view of boxed area shows dual positive TRP63-KRT5 cells (white arrow) and single positive TRP63 cells (white arrowhead) along the basal membrane. Nuclei were counterstained with DAPI. (C) Immunostaining postnatal lung for EpCAM (red), SOX2 (white) and KRT5 (green). Higher power view of boxed area shows single layer of KRT5 cells lining basal membrane. Nuclei were counterstained with DAPI. Br, bronchi; Alv, alveolar region. Scale bars 500 μm (A), 100 μm (B-C). Diagram of 2D air-liquid-interface (ALI) using either postnatal (D) or adult (F) $EpCAM^{+}CD73^{+}$ cells to examine the role of the NOTCH signaling pathway in airway differentiation. (D, F) Representative Z-stacks (upper panel: xy-projection; lower panel zy-projection) through different planes of ALI membranes showing formation of a pseudostratified mucociliary epithelium and impact of DLL4 (10 ng/ml) or DAPT (10 μM) on proximal airway differentiation. Ciliated (β-tubulin, yellow) and goblet cells (MUC5AC, white) can be found at the apical surface

(apical xy plane). Images through the middle xy plane showing club cells (SCGB1A1, purple) together with E-cadherin (Ecad, green) showing changes in cell shape (white arrow). At the basal xy plane, representative images showing basal cells (TRP63, red). (E, G) 3D volume reconstruction of 2D differentiation at ALI to enable visualization basal cells (TRP63, red). n = 3, biological replicates. Scale bars: 500 μm (A), 100 μm (B, C), 30 μm (D, F).

Figure 4. EpCAM$^+$CD73$^+$ cells generate lung organoids recapitulating a mucociliary-secretory cell fate found *in vivo*. (A) Diagram of 3D airway organoid assay using EpCAM$^+$CD73$^+$ cells. (B) Transmission electron microscopy shows ultrastructural analysis of an individual organoid generated from postnatal (top) or adult (bottom) EpCAM$^+$CD73$^+$ cells. n=2, biological replicates. CC, ciliated cell; CR, ciliary rootlet; CZ, contact zone; NSC, nucleus of secretory cell; SC, secretory cell; SG, secretory granules. (C) mRNA levels of genes expressed by airway organoids using RT-qPCR. n=4, biological replicates. Gene expression level in postnatal lung tissue is set at one. (D) Diagram of 3D alveolar organoid assay using EpCAM$^+$CD73$^+$ cells. Phase contrast image (4X) showing generation of organoid structures (right panel). (E) Representative immunofluorescence image of single alveolar organoids. Z stack showing saccule-like features, multicellular organization and lumen formation is highlighted. E-cadherin (red) is used to identify epithelial saccule-like structures and ATII cells stained for proSPC (green, white arrow) are shown. (F) Z-stack of an individual organoid showing KRT5$^+$ cells and the location of ATI cells using the marker HOPX. Magnified Z-axis shows the location of HOPX$^+$ cells facing in towards the lumen of the organoid. Nuclei were counterstained with DAPI (blue). Scale bars: 50 μm (E, F). Two of

three independent experiments are shown. (G) Representative FACS plot of alveolar organoids showing expression of CD73 and HTII-280 gated from $EpCAM^+$ cells (left panel) and isotype for HTII-280 (right panel). (H) Scatter plots show percent of cell subsets within $EpCAM^+$ fraction after subgating for CD73 and HTII-280. (I-J) Scatter plots showing percent of cell subsets within $CD73^+HTII\text{-}280^-$ (I) or $CD73^+HTII\text{-}280^+$ fraction (J) after subgating for CD24 and PDPN. Data are presented as mean ± SD. *P < 0.05; **P < 0.01.

Figure 5. Cystic lung lesions lined with $EpCAM^+CD73^+$ cells. (A-D) Histological analysis of unaffected postnatal lung tissue (A) and congenital lung lesions (B-D). Bottom panels show boxed area at higher power. (E,G,I) Histological analysis (F,H,J) and immunostaining of CPAM (F), Chronic bronchiolitis (H) and CPAM (J) with EpCAM (green) and CD73 (red). Right panels show higher-power view of boxed area and single channels separated without DAPI to highlight co-stained cells. Nuclei are counterstained with DAPI (blue). Br, bronchiole; Alv, alveolar region; Lu, bronchiolar lumen. Scale bars: 500 μm (G), 200 μm (A-D, E, I); 100 μm (F,H,J).

Figure 6. Expansion in $EpCAM^+SCGB1A1^+$ and $EpCAM^+proSPC^+$ cells in distal lung compartment in congenital lesions and other airway abnormalities. Immunostaining shows cystic structures stained with EpCAM (red), SCGB1A1 (white) and proSPC (green) in CPAM (A, C) and lobar emphysema (E). (B,D,F) Right panels of higher-power view of boxed area highlight regions of dual positive EpCAM-SCGB1A1 (yellow arrowhead) or EpCAM-proSPC (yellow arrow) stained cells. Channels are separated to show single stained cells. Br, bronchiole; Alv, alveolar region; Lu,

bronchiolar lumen. Scale bars: 100 μm (A,B,C,E,F); 50 μm (D),. Nuclei are counterstained with DAPI (blue). (G) Increase in $EpCAM^+proSPC^+$ and $EpCAM^+SCGB1A1^+$ cells per total number of $EpCAM^+$ cells in disease compared with healthy area of lung. Data presented as mean ± SD, n = 3, biological replicates.

Figure 7. Re-emergence of $EpCAM^+CD73^+$ cells in NSCLC coincides with increased expression of immune checkpoints and poor prognosis. (A,C) Hematoxylin and Eosin (H&E) staining of tumor specimen. (B) Higher-power view of boxed area from (A) shows dual positive EpCAM (green) and CD73 (Red) staining tumor islands (Tu) (B). (D) Serial section showing the nuclei of dual positive EpCAM-CD73 tumor cells stain for TRP63 (red). Channels are separated to show single stained cells. (E-F) Scatter plots show frequency of $EpCAM^+$ tumor cells enriched for CD73 in (E) LUAD (n = 64) and (F) LUSC (n = 58). (G-H) Scatter plots showing the mean fluorescence intensity (MFI) for PD-L1 and CD47 on $EpCAM^+CD73^+$ cell subset in (F) LUAD (n = 56 - 64) and (H) LUSC (n = 49 - 57). All data determined by flow cytometry. $*P < 0.05$; $***P < 0.0001$; ns, not significant. (I) Representative images of PD-L1 and CD47 expression in serial sections from a single LUAD or LUSC patient. Scale bars: 500 μm (C), 200 μm (A, B,), 100 μm (D), 50 μm (I).

DOI:10.6084/m9.figshare.12488867

Figure 1

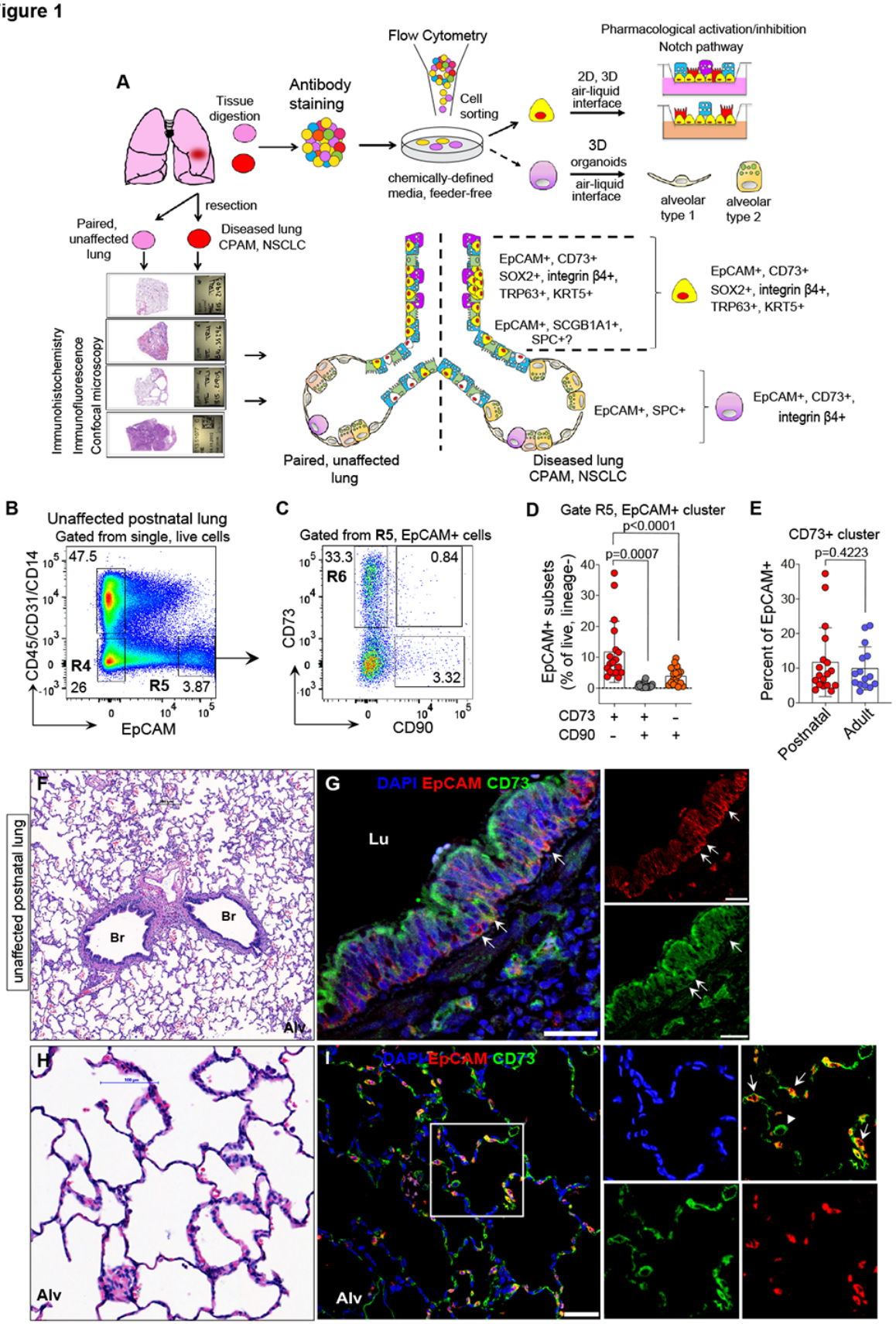

Figure 2

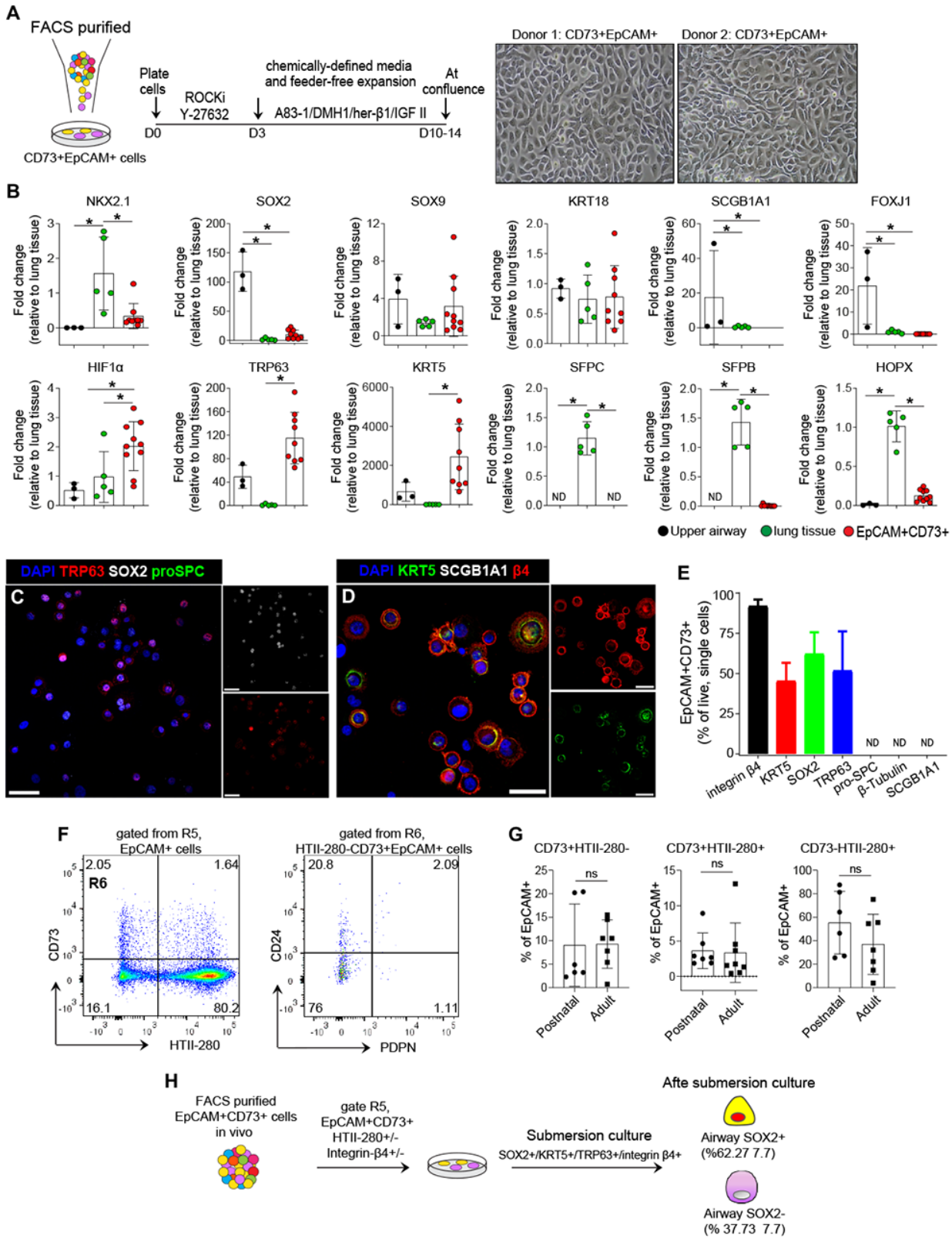

Figure 3

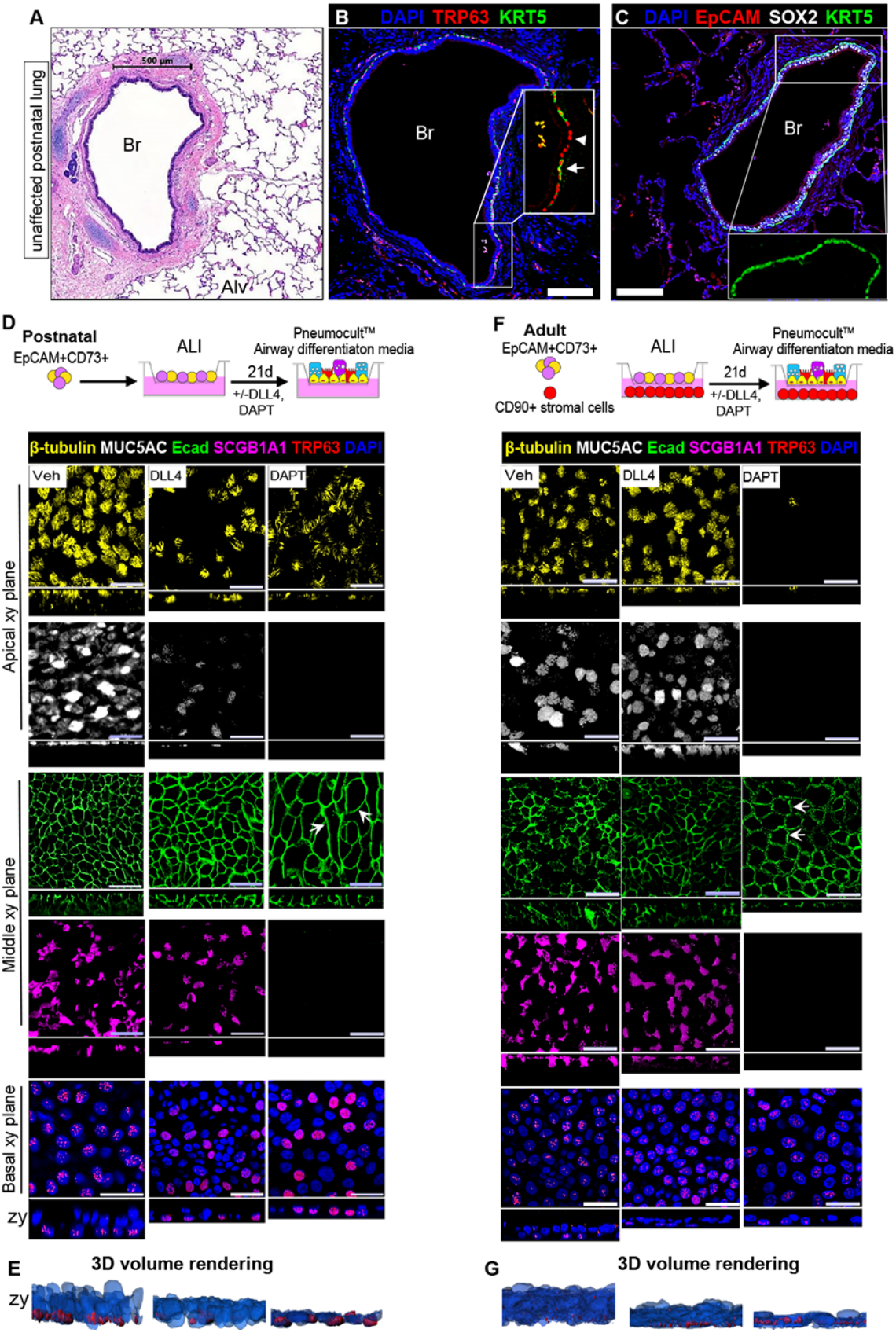

Figure 4

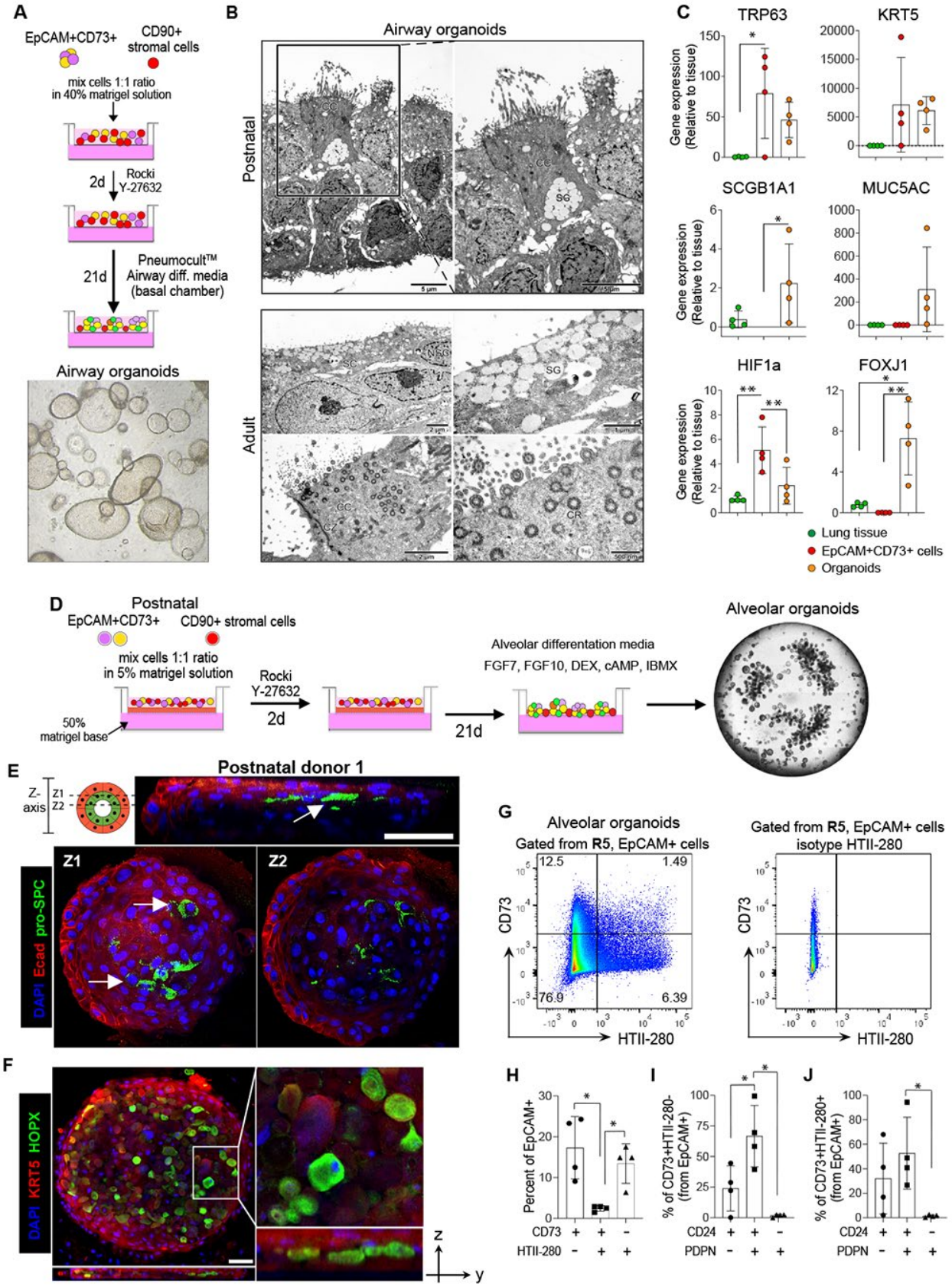

Figure 5

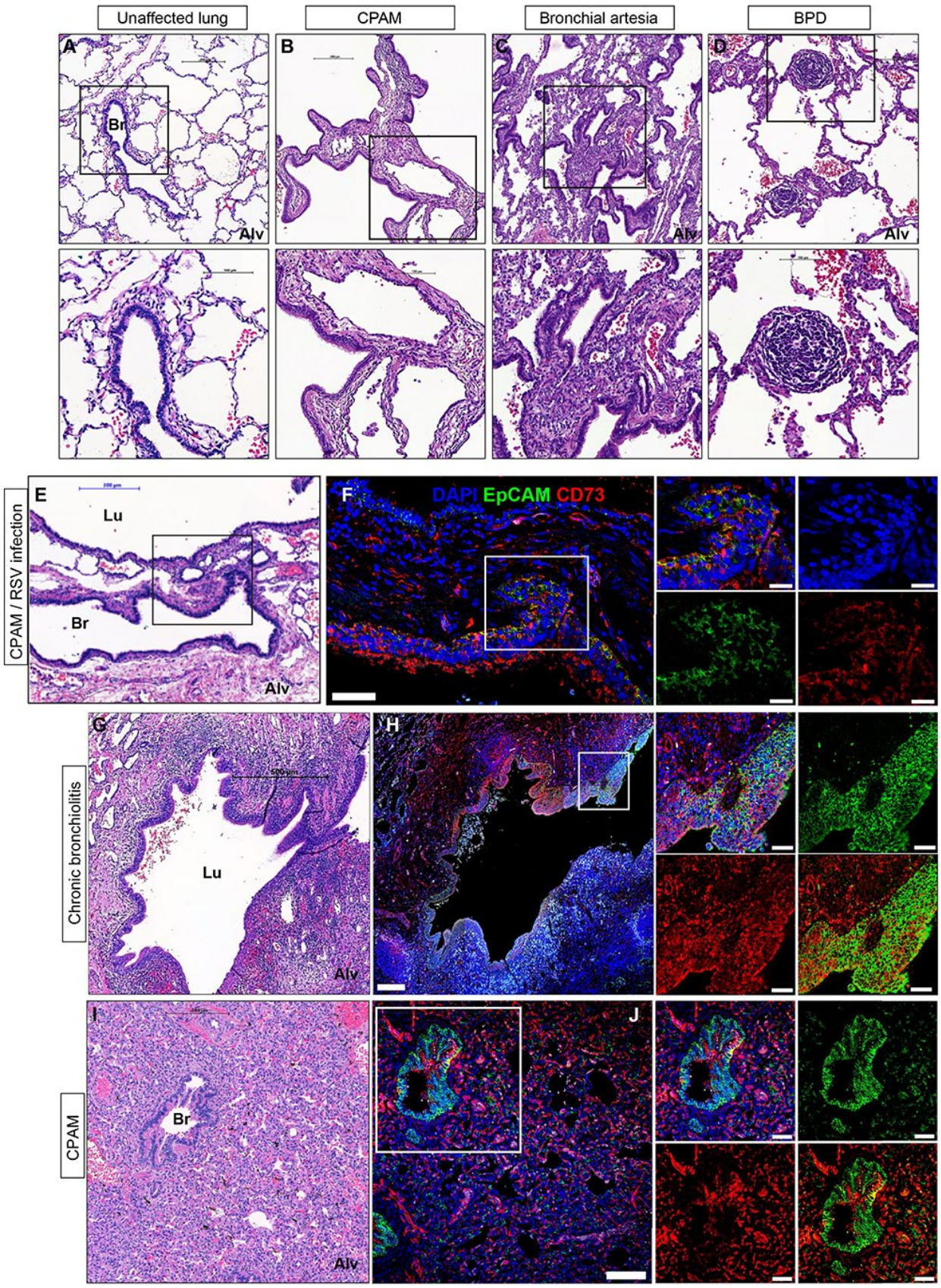

Figure 6

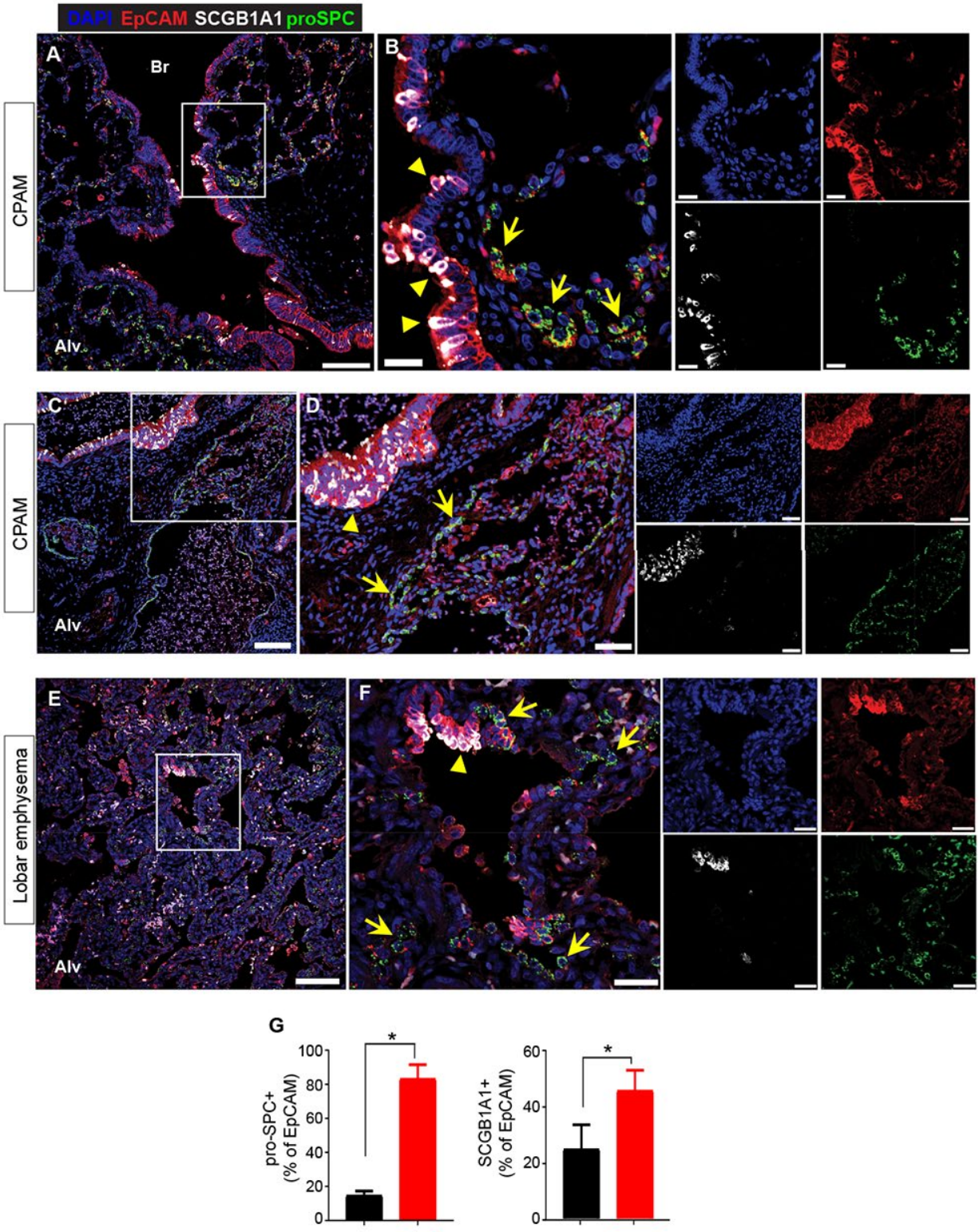

Figure 7